MODELLING EMOTIONAL EFFECTS OF MUSIC: KEY AREAS OF IMPROVEMENT

Tuomas Eerola
University of Jyväskylä, Finland
tuomas.eerola@jyu.fi

ABSTRACT

Modelling emotions perceived in music and induced by music has garnered increased attention during the last five years. The present paper attempts to put together observations of the areas that need attention in order to make progress in the modelling emotional effects of music. These broad areas are divided into theory, data and context, which are reviewed separately. Each area is given an overview in terms of the present state of the art and promising further avenues, and the main limitations are presented. In theory, there are discrepancies in the terminology and justifications for particular emotion models and focus. In data, reliable estimation of high-level musical concepts and data collection and evaluation routines require systematic attention. In context, which is the least developed area of modelling, the primary area of improvement is incorporating musical context (music genres) into the modelling emotions. In a broad sense, better acknowledgement of music consumption and everyday life context, such as the data provided by social media, may offer novel insights into the modelling emotional effects of music.

1. INTRODUCTION

Emotions expressed or induced by music is one of the central aspects in music listening and is one of the main reasons why music appeals to people. The processes involved in emotional communication through music are complicated as they are related to different emotion induction mechanisms, emotion models, expectations, learning, individual differences, and music preferences. The purpose of this paper is to outline the central challenges Music Computing has to face to make advances in emotion modelling in music and outline the necessary steps to ensure forward movement in this field. These challenges can be broadly divided into *theory*, *data* and *context* – the traditional elements of any science – and covered in separate sections of the paper.

In the first section titled *Theory*, issues of theoretical development are discussed. Theory is not perhaps the strongest area of sound and music computing but should not be undervalued since all progress made in the topic requires advances in conceptual and theoretical issues. Issues with emotion models and their prevalence and underlying mechanisms are drawn from recent overviews of the field [1, 2].

In the second section titled *Data*, I refer broadly to representation, collection, processing and interpretation of data. Each of these sub-topics has its own special issues and techniques, many of which have been the focus of studies during the last decade in *Music Information Retrieval* (MIR) and music psychology. The necessity of combining the knowledge and techniques from these separate fields is the central challenge music computing itself has acknowledged (see e.g. roadmap [1]) and the same holds for the field of music and emotion as well.

In the final third section, the *context* of the models and data will be examined. Here, context refers both to the context in which theories and data are supposed to hold and to the contextual constraints provided by the situation, music genre, and individual factors.

2. THEORY

Theoretical issues in music and emotions can be arranged in emotion models, focus, and mechanisms. For modelling, adhering to a particular theoretical framework naturally has vital importance, although the current state of art suggests that the field of music and emotions is not consistent in its use of emotion models, focus, and mechanisms [1,2]. There are terminological differences even within the field of affect sciences (e.g. mood/emotion/feeling) and within the vocabulary sound and music computing studies have adopted from other disciplines (e.g. human-computer interaction, marketing, engineering), whereas certain terms (e.g. mood and emotion) are used interchangeably in some contexts within MIR; these distinctions are important and meaningful when the are communicated across the disciplines. For this reason, I would advocate the conceptual and terminological clarifications drawn by Juslin and Sloboda in the *Handbook of Music and Emotions* [3].

2.1 Emotion models

An important theoretical issue is the notion of how emotions are construed. A plethora of theoretical proposals exist in the psychology of how emotions are divided into *discrete*, *low-* and *high-dimensional* models, and *other* notions for emotions (see Figure 1). According to the discrete emotion model, commonly used in non-musical contexts, all emotions can be derived from a few universal and innate basic emotions such as *fear*, *anger*, *disgust*, *sadness*,

[1] http://mires.eecs.qmul.ac.uk/wiki/index.php/Roadmap

and *happiness* [4]. In music-related studies many of these have been found to be appropriate [5], yet certain emotions have often been replaced by more appropriate ones. For instance, *disgust* is often replaced by *tenderness* or *peacefulness*. Discrete emotion model is commonly utilized in music and emotion studies because it is easy to evaluate in recognition studies, especially with special populations (children, clinical patients, and samples from different cultures) [1].

Low-dimensional models consist of 2 and 3-dimensional models, which propose that all affective states arise from separate independent, affect dimensions. The most common one of these, the two-dimensional circumplex model [6], has one dimension related to *valence* and the other to *arousal*. This particular model has received a great deal of attention in music and emotion studies, despite a number of drawbacks. For instance, it is unable to represent mixed emotions [7], and so several alternative, presumably better, dimensional models have been proposed in which affect the dimensions are chosen differently (e.g., *tension*, *energy*) [8] or by increasing the number of necessary dimensions to three [9, 10]. Recent studies in psychology have generally found formulations other than the valence-arousal dimensions to provide better fit to data [11].

In music, two recent studies of perceived and felt emotions [12, 13] found that the two-dimensional model was found to be a more parsimonious way to represent self-reported ratings of perceived and induced emotions conveyed by film soundtracks. Also, these same studies established that the discrete emotions ratings can be predicted from the ratings of emotion dimensions and vice versa, if the scales and the excerpts are organised in a manner that allows such comparisons.

High-dimensional models of emotions have recently been proposed by Zentner and his colleagues, called *Geneva Emotion Musical Scale* (GEMS) [14], which has from three to nine dimensions of experienced emotions. It has interesting spectrum of terms that emphasize the contemplative, positive and aesthetic nature of music-induced emotions (e.g., *wonder*, *trancendence*, and *nostalgia*). It is worth noting that the GEMS model construction is music-specific and the model construction was carried out with a wide range of participants, and has led to fascinating results on neurophysiological correlates [15]. A direct comparison of low and high-dimensional emotion models in music have, however, suggested that low-dimensional models often suffice to account for the main emotional experiences induced by music [13].

Other theoretical approaches to music and emotion studies include a collection of concepts such as preference, liking, intensity, and also such mood and emotion terms that have been the object of studies recently which have not been connected to theoretical framework. For instance, other types of discrete categories (*passionate, rollicking, humorous, aggressive*) are utilized in *MIREX Audio Mood Classification* task [16]. However, these concepts are not persistently theoretically motivated and may include isolated terms that have little to offer to our understanding of the emotions expressed and induced by music.

There are novel ways to probe which emotion model accounts for the emotions induced and expressed by music. The data provided by social media and online services of music is one such promising source. In the domain of music, social tags describe a variety of information (genre, geography, emotion, opinion, instrumentation, etc.), out of which emotions account for approximately 5% of the most used tags [17]. A number of studies have applied semantic computing to uncover emotion dimensions emerging from the semantic relationships between the tags [18], and some support for the valence-arousal formulation has been found [19]. Such observations have been formalized as *Affective Circumplex Transformation* (ACT) that provides an effective way of predicting the emotional content of music [20].

In sum, a variety of emotion models have been utilized in the sound and music studies and the most common ones have been adopted from psychology, although consensus about their utility has not yet been formed. Also, the models adopted from psychology focus on survival or utilitarian emotions. Music as a pleasurable leisure time activity therefore might be better served with a model that is grounded on terms that are relevant in music-induced emotions such as the ones provided by the GEMS model. Moreover, the emotion models need to be used in the manner consistent with the assumptions build into them. It makes little sense to study valence and arousal using two groups of extreme points within these continuums since the dimensionality cannot be established within such design.

2.2 Emotion focus

Two forms of emotional processes in relation to music can be distinguished – perception and induction of emotions. The first concerns listeners' judgments of emotional characteristics of the music, where listeners characterise the music in emotional terms (e.g., this music is solemn) or what the music may be expressive of (e.g., this music expresses tenderness). Modelling perceived emotions has been the main aim of sound and music computing studies and the most prevalent focus in the field of music and emotions. The latter concerns how music makes listeners feel, also referred to as *felt emotions*. This distinction is not only conceptually plausible, there is also mounting evidence to suggest these two modes of emotional responses can be empirically differentiated [21]. For the field, the problem lies in the often implicit assumption of this division and the induced emotions need to be further validated by indirect measures or psychophysiology. In many instances, we cannot be sure of the distinction. For instance, do emotion related tags or forced-choice selection of facial expressions express felt or perceived emotions?

2.3 Emotion mechanisms

Because the same music can express one emotion and induce another (e.g. cheesy love ballad after a break-up, or a national anthem in a wrong situation), there must be different mechanisms that are responsible for the emotions. The most comprehensive account of the mechanisms to date is the proposal by Juslin and Västfjäll [2], which attempts to

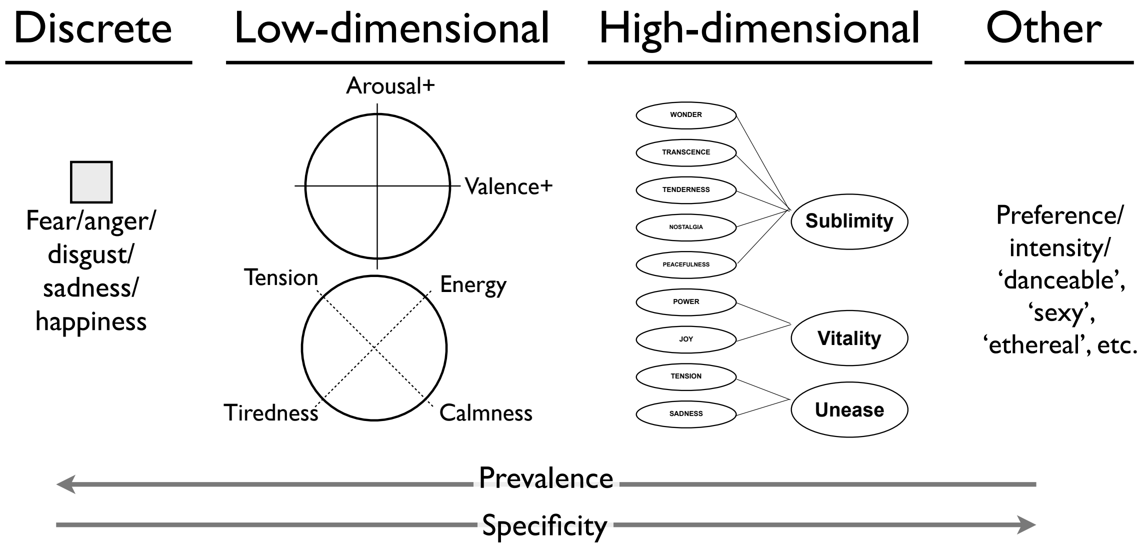

Figure 1. Prevalence and specificity of emotion models applicable to music.

account why music elicits an emotion and why this emotion is of a particular kind. This model, *BRECVEMA* [22], currently consists of eight mechanisms. Each mechanism has distinct response, information focus, possibly brain region, and way of elicitation. However, for sound and music computing, only some of these mechanisms are of central concern. Most past studies have studied *Contagion* mechanism, in which the listener mimics and thus perceives the emotional expression of another being through music, which is also presumed to account for the wide similarity of emotion recognition of music across cultures [23]. *Rhythmic entrainment* is of interest in such cases when the aspects of groove or dancibility have been included in the focus study [24]. Music computing can also attempt to solve the issue of *Musical expectancy*, in which early attempts have already been made [25]. Many other mechanisms are either too limited for application uses or need to be examined in individual settings.

2.4 Epistemological framework

It is also possible to challenge the above-mentioned theoretical issues which emphasise cognitive evaluation of emotions in lieu of other frameworks. Culturally-oriented frameworks would put the emotions in their historical and cultural context [26], and sociological accounts would emphasise how emotions are constructed within particular social groups according to commonly accepted norms constructed in daily lives. The intimate connection of emotions to the body makes embodied cognition a persuasive framework for research [27]. This would emphasise the ecological nature of sound communication and the role of corporeal responses and metaphors in this process. This, in turn, would have implications for what kind of issues will be pursued in emotion research; the process of meaning-generation, empathy, or the underlying neural architecture specialized for mimicry [28]. Finally, application-driven epistemology is something that may generate interesting research in itself, although I would not rank the priority of such research as high.

3. DATA

Sound and music computing is an inherently data-intensive field, and therefore the efforts in music and emotions are directed towards data in its many aspects, specifically (a) representations, (b) processing, (c) collection, and (d) evaluation.

3.1 Data representations

Data representation has specialised in its own areas related to music representations (mostly audio, occasionally midi) and ground-truth representations. In the former, the availability of large amount of good quality audio has widened the scope of studies to include almost any genre, and the number of examples used in studies is only limited by the amount of ground-truth data available for evaluation purposes. This limitation is significant, since availability of audio is meaningless unless it can be connected to listeners' emotions in one way or another. Traditional ground-truth sets contain limited amounts of audio examples carefully assessed by a number of participants in terms of their emotional qualities (self-reports of emotions). Another form of data comes from other measures (indirect, continuous, or physiological) and neural measurements of emotional processing taken during the music listening. These are even more difficult to obtain but have the benefits of being less affected by *demand characteristics*. Moreover, these data representations are more and more supplemented with textual, visual, movement, and social media data, all of which require different tools, algorithms and knowledge from specialized fields. However, combinations of the different data sources is still rare, although most researchers acknowledge the need for multimodal and multiple approaches in emotion research [29].

3.2 Data processing

Data processing borrows from the neighbouring (e.g., computer vision, neuroscience, speech) and technical disciplines (e.g. signal processing). This theme is however, the most advanced one of sound and music computing. However, the processing challenges lie in the realm of temporality of music-induced emotions and synchronisation of physiology and neural responses of the experienced emotions, which all require time-series techniques and behavioural validations. However, these challenges are not unique to music and emotions but pertinent to most neuroscience, physiology and multimedia (movies, particularly) research involved with emotions. Landmark example of how these challenges are solved come from a recent study of music-induced emotions, which correlated the haemodynamic response of the participants with the musical features [30]. Another challenge for data processing concerns the social media data, tags and online meta-data in general, how to obtain semantic structures from such freeform, unconstrained but large datasets [31].

3.3 Musical content estimation

The central limiting factor in predicting emotions from musical content is unreliable estimation of meaningful music-related concepts. Most of the low-level features (e.g. spectral centroid, zero-crossing, or attack slope) have been around for decades but mid to high-level concepts such as tension, mode, harmony and expectancy are demanding to model from audio representation. And this is not only a technical challenge, but rather a conceptual one; high-level concepts require some form of emulation of human perception (e.g. long frame of reference typically modelled with different memory structures, comparisons to typical data structures representing acquired knowledge of regularities in music and so on). Traditionally, there have been two different approaches to this dilemma. An engineering approach applies a combination of low-level features (e.g. MFCCs) and machine learning (e.g. Gaussian Mixture Models or Support Vector Machines) to solve the content problems [32, 33]. Another strategy is to model the perceptual processes faithfully [34], leading in some cases to less efficient models due to emulation of human hearing and all its perceptual constraints (e.g. masking, thresholding, streaming) [35]. Despite the strategy chosen, the need for new and reliable high-level features is strong [36] and reliable measures for syncopation, the degree of "majorness", and expectations are all top priority features that would increase the prediction rates for emotions [37, 38].

Once the features can be estimated reliably, additional steps need to be taken to identify the key features that contribute to emotions. Typically, musical features from an existing music corpus are extracted and mapped into individually rated emotions. The mapping typically takes the form of regression analysis for emotions measurable in scalar terms [39, 40] and emotion categories by means of classification [38]. This approach is correlational because it associates certain features with certain emotions but what it fails to discover is the source of the differences. Another approach is to specifically manipulate musical structure to assess the true effect of these factors to emotions [41]. Unfortunately, the latter approach is time consuming and relatively rare, and typically focuses on few features at a time. Mercifully, combinations of correlational and causal approaches have yielded fairly consistent patterns of results on emotion features in music, summarised by Gabrielsson and Lindström [42].

Because the correlational approach is the most common and offers the largest sets of data, it is important to consider the feature selection before the construction of the model. Elsewhere, I have suggested four stages for this process [43]; (a) theoretically select plausible features, (b) validate the chosen features, (c) optimise the chosen features, and (d) evaluate the predictive capacity of the model. Theoretical selection is justified to eliminate dozens of technically possible features that may just increase noise. In the next step, the researcher should verify that the features are reliable and provide relevant information using a separate ground-truth dataset. In the third step, exploration of the independence of the features is useful in order to trim the feature set into separate, independent and preferably orthogonal entities using data reduction techniques. These steps decrease the danger of over-fitting and facilitate the interpretation of the subsequent models.

3.4 Data collection, evaluation and access

Finally, the data is as good as the collection and evaluation procedures allow it to be. In sound and music computing, rigorous data collection procedures are not always adhered to due to emphasis on algorithm development or data modelling, or in some cases, the researchers may not always have the expertise to follow the methodological requisites perfected in the behavioural sciences (e.g. psychology). Participant background descriptions (music preference and musical sophistication indices), and outlier screening, inter-rater reliability, and general replicability are often neglected in the data evaluation procedures in small-scale behavioural studies. Despite these traditional concerns, there are new innovative ways of getting participant data. Online games have been found to be a good way in obtaining mood ratings [44], crowd-sourcing platforms (e.g. *Amazon Mechanical Turk*), and large-scale online questionnaires that have certain practical limitations (sound setup, situation, listener background) but the large participant amount is assumed to compensate for these drawbacks. Another data collection issue is the annotation. Expert annotations are expensive and laborious, and crowd-sourced annotations may in some situations lead to equally coherent results [45]. Whether the data obtained from certain social online music services (e.g. *last.fm*, *Spotify* see *Million Song Dataset* [46]) can be harnessed to tackle the fundamental issues related to music and emotions, still remains to be seen but the results so far are promising in non-music related domains [47] and in music [20, 31].

Also, the modelled data needs to be assessed in a rigorous fashion. Whereas the studies adhering to psychology standards typically collect and evaluate the data properly, they often produce a final model that accounts for the handful of excerpts that are also the ones used to train the model in the study and no cross-validation and prediction with external datasets are used. Fortunately, sound and music studies normally pay attention to these issues and some researchers have taken the cross-validation steps particularly seriously [37, 38].

Finally, the effectiveness of the music and emotion research would be increased by establishing common repositories for open data-sharing (stimuli, features, evaluations, and protocols) and therefore facilitating replicability of the studies [48]. There are already shared tools (toolboxes such as *Marsyas*, *Sonic Visualiser*, and *MIR toobox* for musical feature extraction) and platforms for data sharing [49], and also possibilities of organising all this in an open and attributable manner (e.g. http://thedata.org/). In certain cases, this is routinely done [12,50] but the strength of sound and music computing is not fully capitalised before many different datasets are openly available.

4. CONTEXT

Theories and data only operate in the context in which they have been defined. In music psychology, the context of music and emotion studies have mainly been in Western art music and highly Western educated listeners in particularly restricted situations (concerts or laboratory setting), judging from the frequency of music genres, situations and participants utilised in the past ten years [1]. In sound and music computing, the context is more consumption oriented, that is, more studies utilising pop music and everyday listening situations and therefore closer to current music consumption habits [51]. However, context is much more; here broadly divided into socio-cultural, musical, individual and listening context.

4.1 Socio-cultural context

For modelling emotions in music, the cultural context is certainly the largest open issue that not only divides listeners in Western countries according to geographical areas and age groups, but to broad cultural differences across the globe. Few cross-cultural studies of emotion recognition have been conducted which explore the topic using music excerpts and listeners from multiple cultures [23, 52]. Fortunately, in sound and music computing, this issue has been acknowledged for some time now [53, 54] and datasets and applications of existing techniques to novel musical materials are at least applied to non-Western music collections [55]. This recent tendency has also highlighted the need for further development of musical feature extraction due to challenges offered by non-Western tuning systems and instruments. Within a culture, there are wide differences in musical practices, consumption habits, and meanings associated with music between different social and age groups. These socio-cultural differences have not received the attention they deserve, although they are known to have wide impact on music choices and emotions induced by music.

4.2 Musical context

As a smaller subset of the cultural context, the musical context – music genre, lyrics and videos – brings tangible differences for modelling emotions in music. Just consider genre differences; what is recognised as tender in piano music of *late romantic era*, probably does not have relevance in *gothic metal*, and happiness in *pop* may not be equivalent either as a concept or musical term in *electronica*. Recently, sobering results from the generalisibility of simple emotion predictions of valence and arousal across music genres was obtained [37]. According to the results, emotional valence did not transfer across genres although arousal did. In a small-scale study, the same musical features have been shown to operate differently if the underlying context is changed [56]. When the large materials provided by social media tags is harnessed for emotions in music, it has been found that genre information is able to bring significant improvements on model predictions [20]. For modelling emotions in music, the role of genre seems to be of utmost importance.

4.3 Individual context

With the context I also refer to individual differences such as personality, motivation and self-esteem, which all bring about significant differences between listeners. Such personality traits as neuroticism and extraversion are linked with negative and positive emotionality, leading to differences in music-induced emotions as well [57]. It is also known that specific personality traits, such as *openness to experience*, are linked with music-induced chills [58]. For modelling emotions in music, the individual differences have less important roles than say, music genre, but nevertheless, there is now a trend to incorporate the individuality of the user when creating personalised recommendation systems for music [59].

4.4 Listening context

A host of situational factors affect emotions induced by music. From everyday music listening studies [60] we know that differences in the listening contexts – whether at home, at a laboratory, on public transport, with friends, etc. – has a strong influence on what emotions are likely to be experienced. For instance, it is known that emotional episodes linked with music are most common at home and at evening, and occur during music listening, social interaction, or relaxation, working and watching movies or TV. These situational and social factors are challenging to incorporate into the emotion modelling. However, the contextual information provided by the situation is something that at least needs to be acknowledged in modelling emotions in music, even if it states that these results generally hold for people listening to music alone in laboratory conditions.

5. CONCLUSIONS

Significant advances in all areas of modelling emotional effects of music have been made during the last decade.

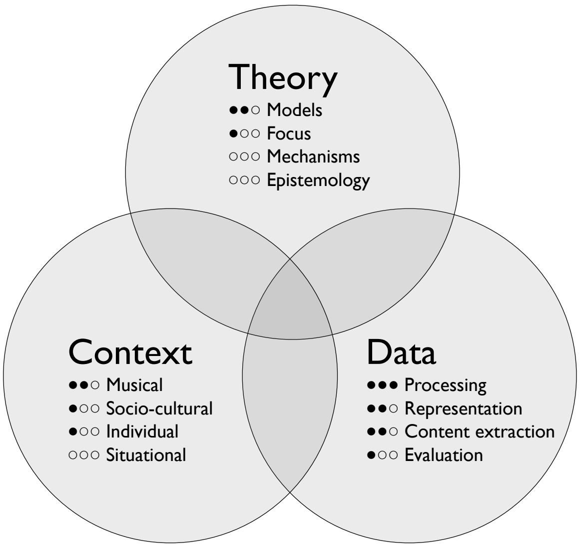

Figure 2. Key areas and their current status in modelling emotions in music (filled circles indicate advanced status).

Figure 2 emphasizes how the areas overlap and need to be developed in tandem. Figure also summarizes the current progress of the important areas. Those areas that are particularly well developed are ranked high (shown with small black indicators) and those key areas that require further attention can be summarized:

- commitment to emotion focus and mechanisms
- estimation of high-level music content
- robust evaluation procedures
- open data sharing conventions
- everyday listening (e.g. data and functions)
- sensitivity to musical context (e.g. genres)

These key areas of attention have been the subject of some studies detailed in earlier sections, but the progress in them is still limited. In the theoretical domain – which has lesser status in sound and music computing – future studies should adopt critical outlook to emotion models, focus and underlying theoretical assumptions. In the domain of data, cross-validation, appropriate behavioural data collection practices, creation of ways to measure high-level concepts from audio, and making all the efforts transparent by sharing the code and the data would greatly speed up the progress made in the field. Any advances in context-related issues would be a significant improvement, but to create better models of emotional effects of music, taking into account inherent differences in emotional values and functions of different music genres would provide the most imminent benefits.

6. REFERENCES

[1] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models and stimuli," *Music Perception*, vol. 30, no. 3, pp. 307–340, 2012.

[2] P. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and Brain Sciences*, vol. 31, no. 05, pp. 559–575, 2008.

[3] P. N. Juslin and J. A. Sloboda, *Handbook of Music and Emotion*. Boston, MA: Oxford University Press, 2010, ch. Introduction: Aims, organization, and terminology, pp. 3–12.

[4] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, pp. 169–200, 1992.

[5] P. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[6] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[7] P. G. Hunter, E. G. Schellenberg, and U. Schimmack, "Mixed affective responses to music with conflicting cues," *Cognition & Emotion*, vol. 22, no. 2, pp. 327–352, 2008.

[8] R. E. Thayer, *The Biopsychology of Mood and Arousal.* Oxford University Press, New York, USA, 1989.

[9] U. Schimmack and A. Grob, "Dimensional models of core affect: A quantitative comparison by means of structural equation modeling," *European Journal of Personality*, vol. 14, no. 4, pp. 325–345, 2000.

[10] H. Lövheim, "A new three-dimensional model for emotions and monoamine neurotransmitters," *Medical Hypotheses*, vol. 78, no. 2, pp. 341–348, 2012.

[11] D. C. Rubin and J. M. Talarico, "A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words," *Memory*, vol. 17, no. 8, pp. 802–808, 2009.

[12] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011.

[13] J. K. Vuoskoski and T. Eerola, "Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences," *Musicae Scientiae*, vol. 15, no. 2, pp. 159–173, 2011.

[14] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Differentiation, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[15] W. Trost, T. Ethofer, M. Zentner, and P. Vuilleumier, "Mapping aesthetic musical emotions in the brain," *Cerebral Cortex*, vol. 22, no. 12, pp. 2769–2783, 2012.

[16] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann, "The 2007 MIREX audio mood classification task: Lessons learned," in *Proceedings of the 9th International Conference on Music Information Retrieval*, 2008, pp. 462–467.

[17] P. Lamere, "Social tagging and music information retrieval," *Journal of New Music Research*, vol. 37, no. 2, pp. 101–114, 2008.

[18] M. Levy and M. Sandler, "A semantic space for music derived from social tags," in *Proceedings of 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

[19] C. Laurier, M. Sordo, J. Serra, and P. Herrera, "Music mood representations from social tags," in *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR)*, 2009, pp. 381–86.

[20] P. Saari and T. Eerola, "Semantic computing of moods based on tags in social media of music," *IEEE Transactions on Knowledge and Data Engineering*, manuscript submitted for publication available at http://arxiv.org/, 2013.

[21] P. Evans and E. Schubert, "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, vol. 12, no. 1, pp. 75–99, 2008.

[22] P. N. Juslin, "From everyday emotions to aesthetic emotions: Toward a unified theory of musical emotions," *Physics of Life Reviews*, in press.

[23] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. D. Friederici, and S. Koelsch, "Universal recognition of three basic emotions in music," *Current Biology*, vol. 19, no. 7, pp. 573–576, 2009.

[24] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambó, E. Gómez, P. Herrera *et al.*, "Semantic audio content-based music recommendation and visualization based on user preference examples," *Information Processing & Management*, vol. 49, no. 1, pp. 13–33, 2012.

[25] M. M. Farbood, "A parametric, temporal model of musical tension," *Music Perception: An Interdisciplinary Journal*, vol. 29, no. 4, pp. 387–428, 2012.

[26] L. Kramer, *Music as cultural practice, 1800-1900*. Berkeley, US: University of California Press, 1990.

[27] M. Maiese, *Embodiment, emotion, and cognition*. New York, US: Palgrave, 2011.

[28] I. Molnar-Szakacs and K. Overy, "Music and mirror neurons: from motion to'e'motion," *Social cognitive and affective neuroscience*, vol. 1, no. 3, pp. 235–241, 2006.

[29] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner *et al.*, "The humaine database: addressing the collection and annotation of naturalistic and induced emotional data," in *Affective computing and intelligent interaction*. Springer, 2007, pp. 488–500.

[30] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *NeuroImage*, vol. 59, no. 4, pp. 3677–3689, 2012.

[31] M. Levy and M. Sandler, "Learning latent semantic models for music from social tags," *Journal of New Music Research*, vol. 37, no. 2, pp. 137–150, 2008.

[32] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE transactions on*, vol. 10, no. 5, pp. 293–302, 2002.

[33] Q. Claire and R. D. King, "Machine learning as an objective approach to understanding music," in *New Frontiers in Mining Complex Patterns*. Springer, 2013, pp. 64–78.

[34] A. Novello, S. van de Par, M. M. McKinney, and A. Kohlrausch, "Algorithmic prediction of inter-song similarity in western popular music," *Journal of New Music Research*, no. ahead-of-print, pp. 1–19, 2013.

[35] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *Proc. ISMIR*, 2005, pp. 34–41.

[36] K. Markov and T. Matsui, "High level feature extraction for the self-taught learning algorithm," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, pp. 1–11, 2013.

[37] T. Eerola, "Are the emotions expressed in music genre-specific? an audio-based evaluation of datasets spanning classical, film, pop and mixed genres," *Journal of New Music Research*, vol. 40, no. 4, pp. 349–366, 2011.

[38] P. Saari, T. Eerola, and O. Lartillot, "Generalizability and simplicity as criteria in feature selection: Application to mood classification in music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1802–1812, 2011.

[39] T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models," in *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR 2009)*, K. Hirata and G. Tzanetakis, Eds. Dagstuhl, Germany: International Society for Music Information Retrieval, 2009, pp. 621–626.

[40] Y. Yang, Y. Lin, Y. Su, and H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio Speech and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.

[41] P. N. Juslin and E. Lindström, "Musical expression of emotions: Modelling listeners' judgements of composed and performed features," *Music Analysis*, vol. 29, no. 1-3, pp. 334–364, 2010.

[42] A. Gabrielsson and E. Lindström, "The role of structure in the musical expression of emotions," *Handbook of music and emotion: Theory, research, applications*, pp. 367–400, 2010.

[43] T. Eerola, "Modeling listeners' emotional response to music," *Topics in Cognitive Science*, vol. 4, no. 4, pp. 607–624, 2012.

[44] Y. E. Kim, E. Schmidt, and L. Emelle, "Moodswings: A collaborative game for music mood label collection," in *Proceedings of the International Symposium on Music Information Retrieval*, 2008, pp. 231–236.

[45] P. Saari, M. Barthet, G. Fazekas, T. Eerola, and M. Sandler, "Semantic models of mood expressed by music: Comparison between crowd-sourced and curated editorial annotations," in *IEEE International Conference on Multimedia and Expo (ICME 2013): International Workshop on Affective Analysis in Multimedia (AAM)*, In press 2013.

[46] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

[47] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "Mood sensing from social media texts and its applications," *Knowledge and Information Systems*, pp. 1–36, 2013.

[48] R. Mayer, A. Rauber, and S. B. Austria, "Towards time-resilient mir processes," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2012, pp. 337–342.

[49] K. West, A. Kumar, A. Shirk, G. Zhu, J. S. Downie, A. Ehmann, and M. Bay, "The networked environment for music analysis (nema)," in *Services (SERVICES-1), 2010 6th World Congress on*. IEEE, 2010, pp. 314–317.

[50] J. Skowronek, M. McKinney, and S. ven de Par, "Ground-truth for automatic music mood classification," in *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, 2006, pp. 395–396.

[51] T. Lidy and P. van der Linden, "Report on 3rd chorus+ think-tank: Think-tank on the future of music search, access and consumption, midem 2011," CHORUS+ European Coordination Action on Audiovisual Search, Cannes, France, Tech. Rep., March 15 2011.

[52] P. Laukka, T. Eerola, N. S. Thingujam, T. Yamasaki, and G. Beller, "Universal and culture-specific factors in the recognition and performance of musical emotions," *Emotion*, in press.

[53] T. Lidy, C. N. Silla Jr, O. Cornelis, F. Gouyon, A. Rauber, C. A. Kaestner, and A. L. Koerich, "On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections," *Signal Processing*, vol. 90, no. 4, pp. 1032–1048, 2010.

[54] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright, "Computational ethnomusicology," *Journal of interdisciplinary music studies*, vol. 1, no. 2, pp. 1–24, 2007.

[55] Y.-H. Yang and X. Hu, "Cross-cultural music mood classification: A comparison of english and chinese songs," in *Proc. ISMIR*, 2012.

[56] T. Eerola, "Analysing emotions in schubert's erlkönig: A computational approach," *Music Analysis*, vol. 29, no. 1-3, pp. 214–233, 2010.

[57] J. K. Vuoskoski and T. Eerola, "The role of mood and personality in the perception of emotions represented by music," *Cortex*, vol. 47, no. 9, pp. 1099–1106, 2011.

[58] E. C. Nusbaum and P. J. Silvia, "Shivers and timbres personality and the experience of chills from music," *Social Psychological and Personality Science*, vol. 2, no. 2, pp. 199–204, 2011.

[59] A. S. Lampropoulos, P. S. Lampropoulou, and G. A. Tsihrintzis, "A cascade-hybrid music recommender system for mobile services based on musical genre classification and personality diagnosis," *Multimedia Tools and Applications*, vol. 59, no. 1, pp. 241–258, 2012.

[60] P. Juslin, S. Liljeström, D. Västfjäll, G. Barradas, and A. Silva, "An experience sampling study of emotional reactions to music: Listener, music, and situation." *Emotion*, vol. 8, no. 5, pp. 668–683, 2008.